Grid Technology as a CyberInfrastructure for Earth Science Applications

Thomas H. Hinke, Ph.D.

NASA Ames Research Center

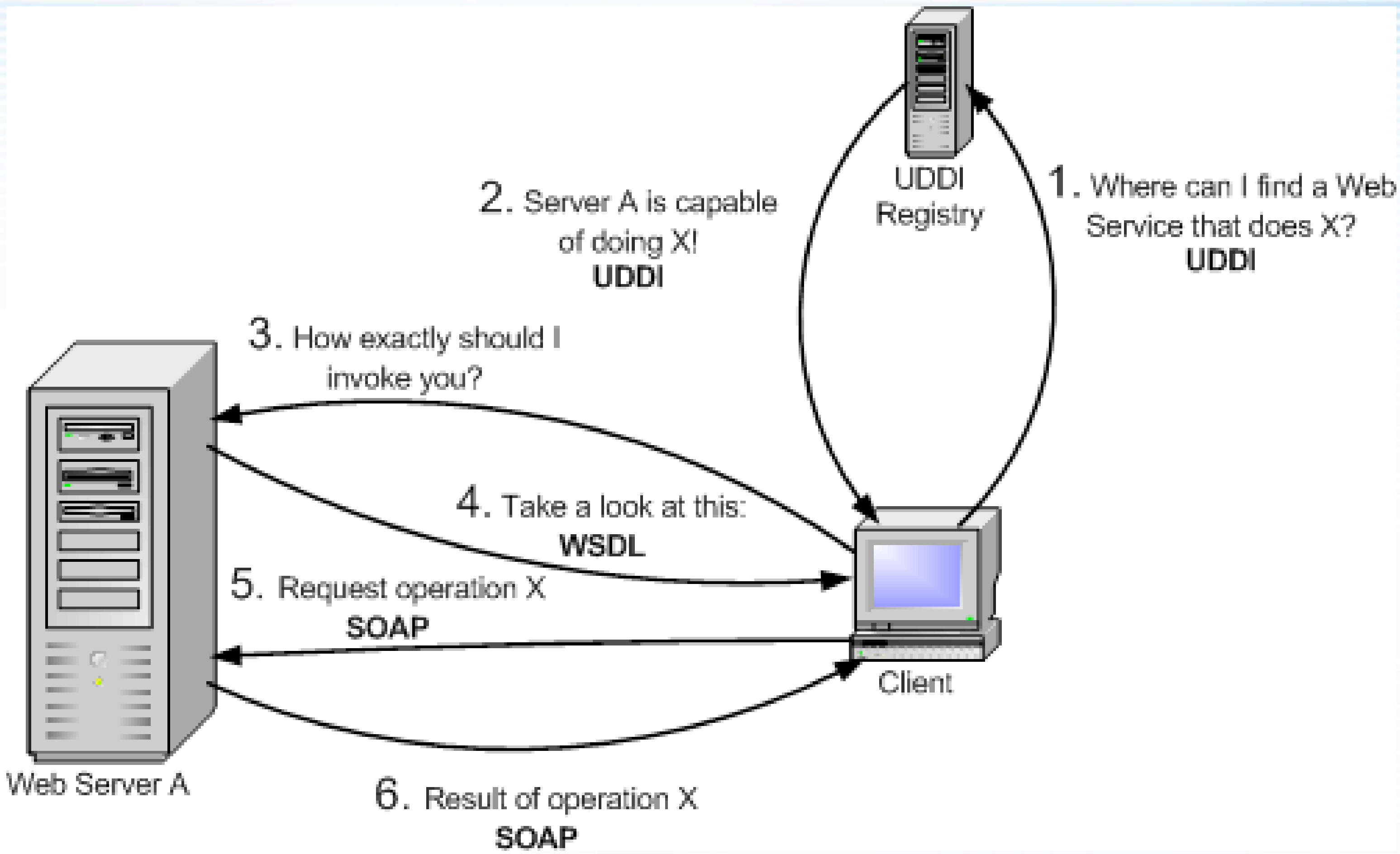
Outline

- What are grid services and their importance to Earth Science
- How can grid services support Earth Science applications
- NASA's role in grids and grid services
- Roadmap to an Earth Science oriented service based cyberInfrastructure

Evolution of Grid Services

- Out of the grid community comes technology that
  - Provides secure single-sign-on
  - Provides seamless access to various types of resources
    - Computational resources
    - Data archives
    - Scientific Sensors/Instruments
- Out of the web service community came technology that
  - Provides a service oriented architecture
    - Data sources and applications expressed as self-describing services
  - Uses commercially available web service technology
  - Allows more complex services to be constructed from simpler service.
- Web and Grid services have been combined
  - Standardization underway

Web/Grid Services Support Software
Access To Services and Data



Graphic from Borja Sotomayor http://www.casa-sotomayor.net/gt3-tutorial/index.html

Grid Services Provide A CyberInfrastructure for Applications

- Grid Services are based on commercial Web services but support grid security
- They support remote access to data by software, not human access as is the goal of web page technology
    - Web has to do with human access to data using web browsers
    - Web services have to do with software access to data
- They use the Web Services Description Language (WSDL) to describe their interface.
- They are intended for loosely coupled distributed systems
- Standardization efforts are underway
    - OGSA (Open Grid Services Architecture) by the Global Grid Forum (GGF) for the higher level core services
    - WS-RF (Web Service Resource Framework) by OASIS (Organization for the Advancement of Structured Information Standards) based on earlier work on GGF Open Grid Services Infrastructure (OGSI) standard for the lower level services

Benefits of Grid Services to the Earth Science Community

- Provides re-usable building blocks for other services and applications
- Reduces development costs for future Earth Science system development, replacement and upgrades
- Provides an architecture to allow improved services to be easily integrated into a system (since WSDL description describes how to access service)
- Allows users and applications to easily use data and computational resources, irrespective of their location within the NASA community

NASA ADVANCED SUPERCOMPUTING

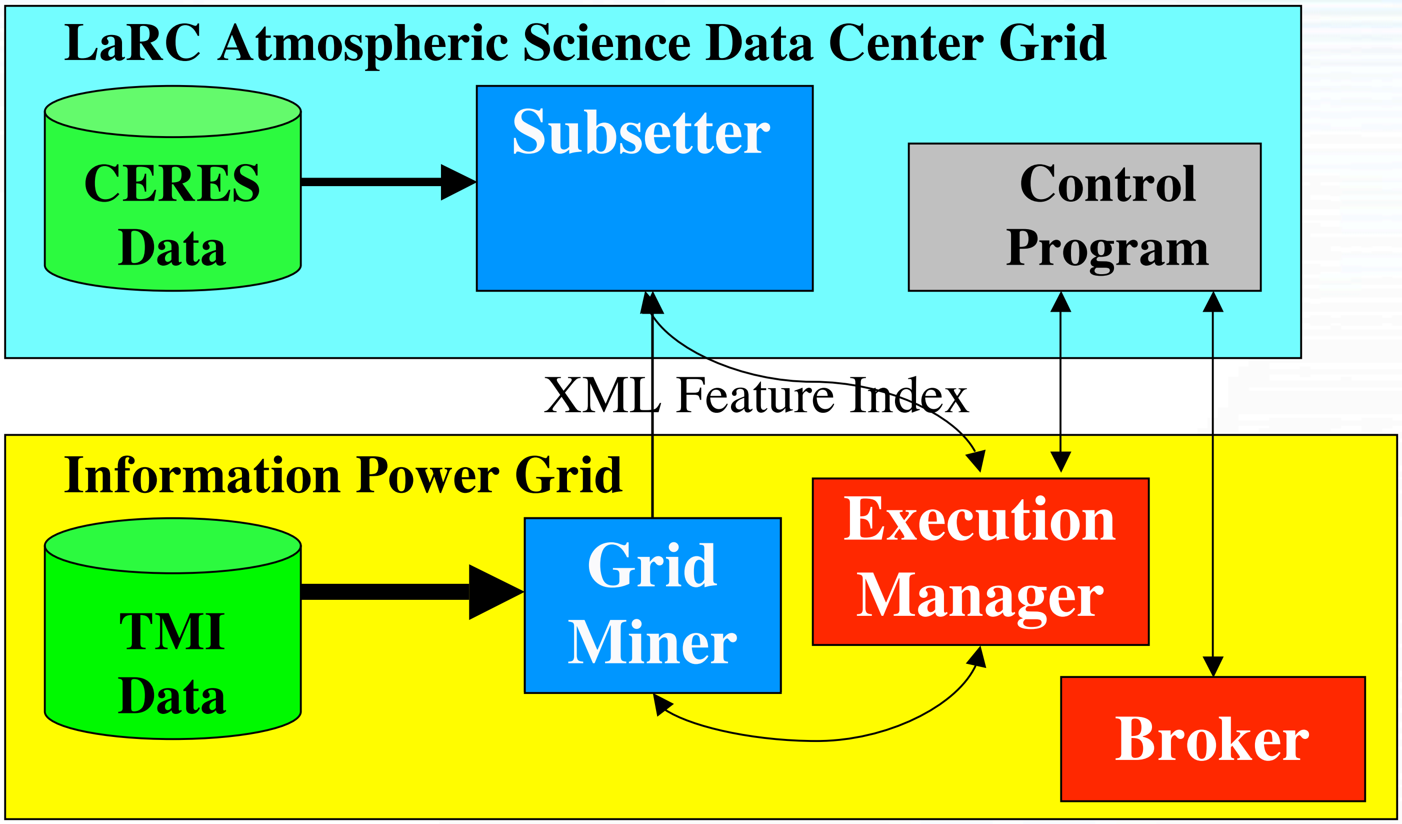
Outline

- What are grid services and their importance to Earth Science
- How can grid services support Earth Science applications
- NASA's role in grids and grid services
- Roadmap to an Earth Science oriented service based cyberInfrastructure

Grid Services Can Support Earth Science in at Least Four Ways

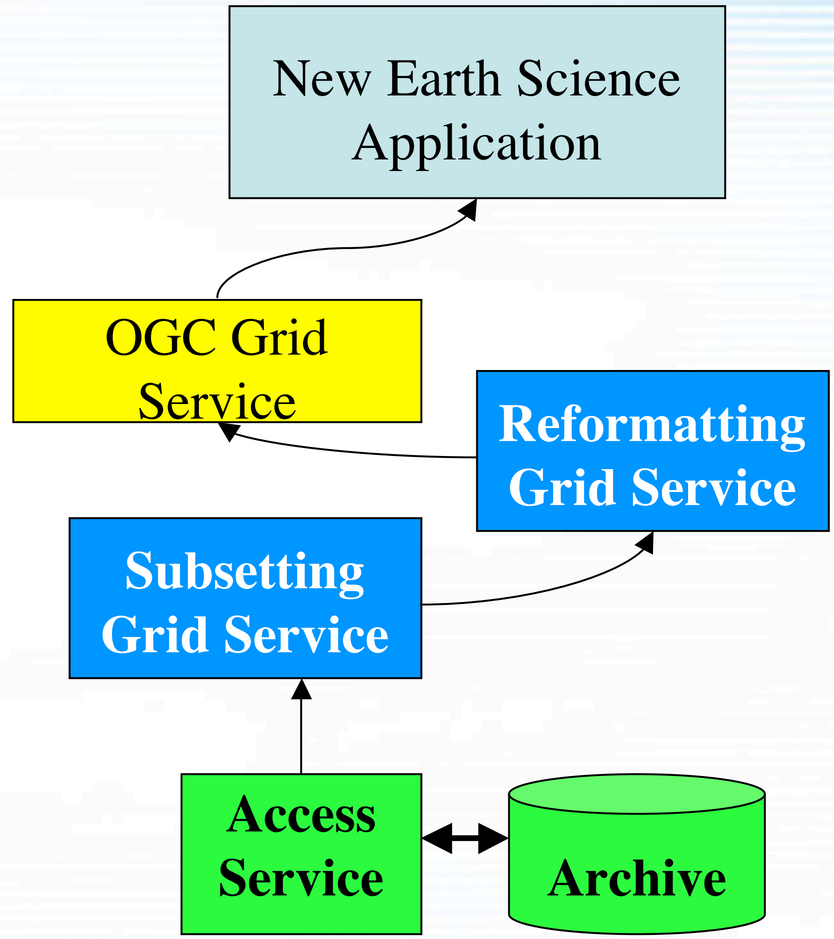
- As core services needed for any distributed system
- As general purpose Earth Science Services that can be reused by many different Earth Science applications
- As services that required high-end processing
- As as a way of integrating space-based data with Earth-based processing
- As a way of sharing expertise among the international community

Core Services Facilitate Application Development



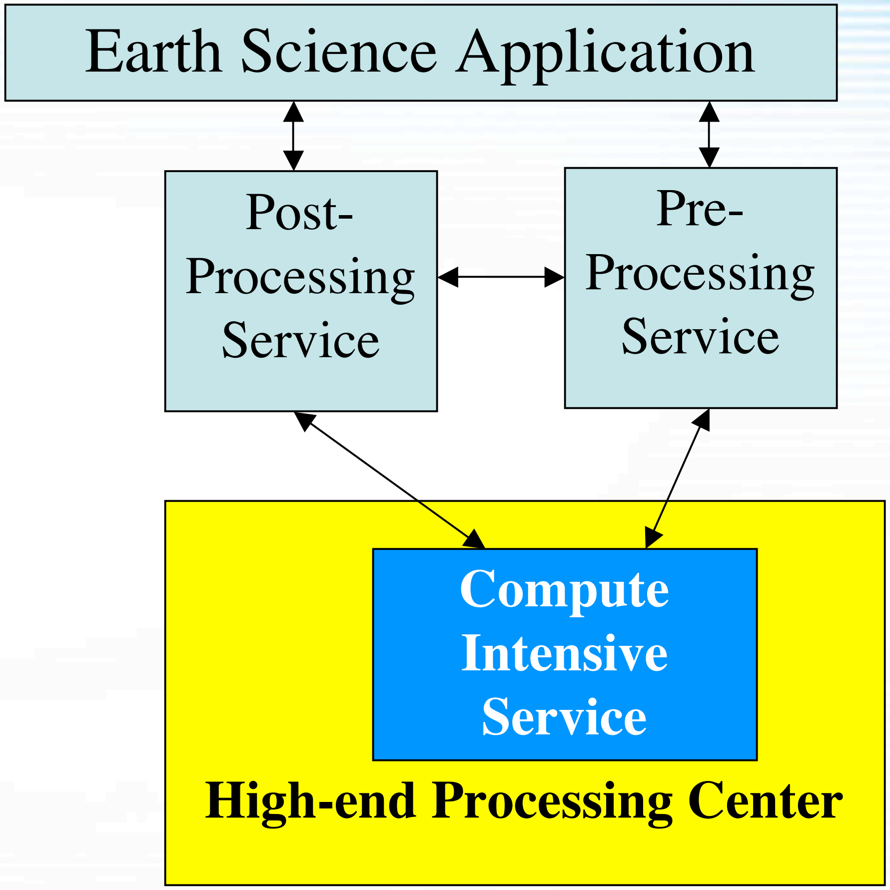
Earth Science Services Can Be Shared and Used to Constructed New Services

- Widely needed Earth Science functions can be configured as services:
  - Data Archive Access
  - Subsetter, Reformatter
  - Open GIS (Geographical Information System) Consortium (OGC) compliant services
- Grid supports movement of
  - Data to service
  - Service to data repository
- Service could be hosted/maintained by those with expertise in the service, but shared with the wider Earth Science Community



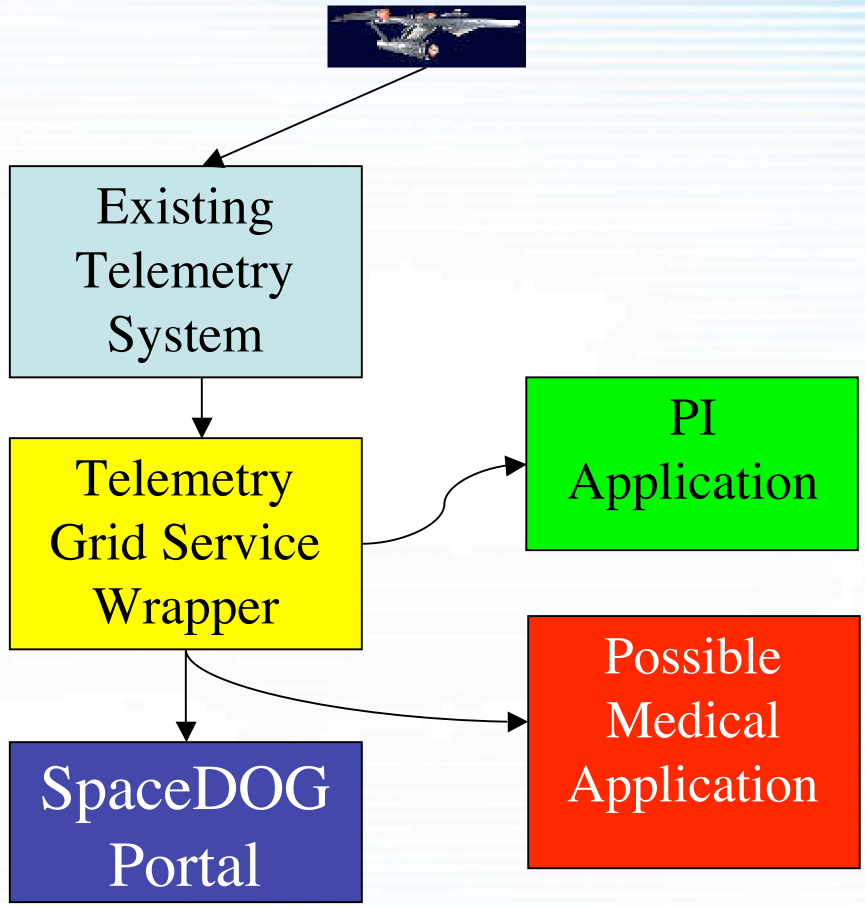
Some Earth Science Applications May Required High End Processors

- Service could be maintained by organization with expertise in service
- Service could be hosted at location with high end processors
- High speed networks would deliver data to processors



Project to Provide Service-Based Access Telemetry Data

- In collaboration with MSFC and JSC developing telemetry service
- One part of Space Development & Operations Grid (SpaceDog) POC: Bob Bradford at MSFC.
- Potential benefit: Reduce development costs for future space flight system development, replacement and upgrades
- Could extend grid to space, and include satellite sensors as elements of the grid, accessed through a grid service



Committee on Earth Observation Satellites (CEOS) Grid Testbed: First Step Toward International Sharing

Objectives: The CEOS GRID Task Team will develop technical requirements and identify GRID technologies and services to be implemented in testbed locations

Approach:

- ARC and others supported CEOS Grid Workshop in April 2002
- CEOS Grid Testbed participants began work in 2002 and have grown since then
- ARC provided grid consulting as well as host and user grid certificates
- Various CEOS participants are developing grid applications

Participants:

- EOSDIS & George Mason University (GMU)
- European Space Agency (ESA)
- DutchSpace
- NOAA Operational Model Archive & Distribution System (NOMADS)
- University of Alabama – Huntsville (UAH)
- United States Geological Survey – EROS Data Center (EDC)
- NASA Advanced Data Grid (ADG)
- China Spatial Information Grid (SIG)
- ARC (supplying host and user certificates and grid consulting)

Benefits: Vehicle for exploring the advantage of Grid technology for Earth Science

Outline

- What are grid services and their importance to Earth Science
- How can grid services support Earth Science applications
- NASA's role in grids and grid services
- Roadmap to an Earth Science oriented service based cyberInfrastructure

Current Status of Grids at NASA

- NASA Ames Research Center (ARC) has developed the Information Power Grid that currently encompasses computational resources at
  - Ames Research Center, Glenn Research Center, Langley Research Center, JPL, with planning underway for Goddard Space Flight Center
- NASA is developing grid service technology to increase the intelligence of the grid as a cyber infrastructure
- Since NASA may ultimately have many grids NASA ARC
  - Has provided technical support to other NASA and NASA-related organizations (e.g., CEOS) that are interested in deploying grid technology
  - Interested in services that may need to support of high-end computing resources

NASA Focus: Increasing Intelligence of Grid Processing and Data Handling

- Resource discovery service through brokers that select the "best" set of resources based on user requirements
- Execution manager service that can autonomously manage the user's job as it moves through stages of grid processing
- Naturalization service that automatically tailors the processing environment on grid resources
- Dynamic access service that permits users to instantly, but with proper accountability, access needed computational resources across administrative boundaries, without having pre-established accounts on these machines
- Data discovery through distributed, grid-accessible metadata catalogs

NASA Grid Development

Application and User Oriented Services
- Portal Services
- Workflow Specification
- Query Structuring

Execution Management Services
- Workflow Management
- Resource Broker
- Naturalization Services
- Remote Execution
- Task Graph Execution

Data Management Services
- Replica Management
- Meta-data Catalogues
- Data Movement

Grid Management Services
- Distributed Information Management
- Dynamic User Access Management
- Event Management
- Test & Monitoring Service
- Information Service
- Security Infrastructure

Application Integration Processes and Tools

- ARC in-house R & D Efforts
- ARC partnership with external organizations
- External Technologies

Compute & Data Resources

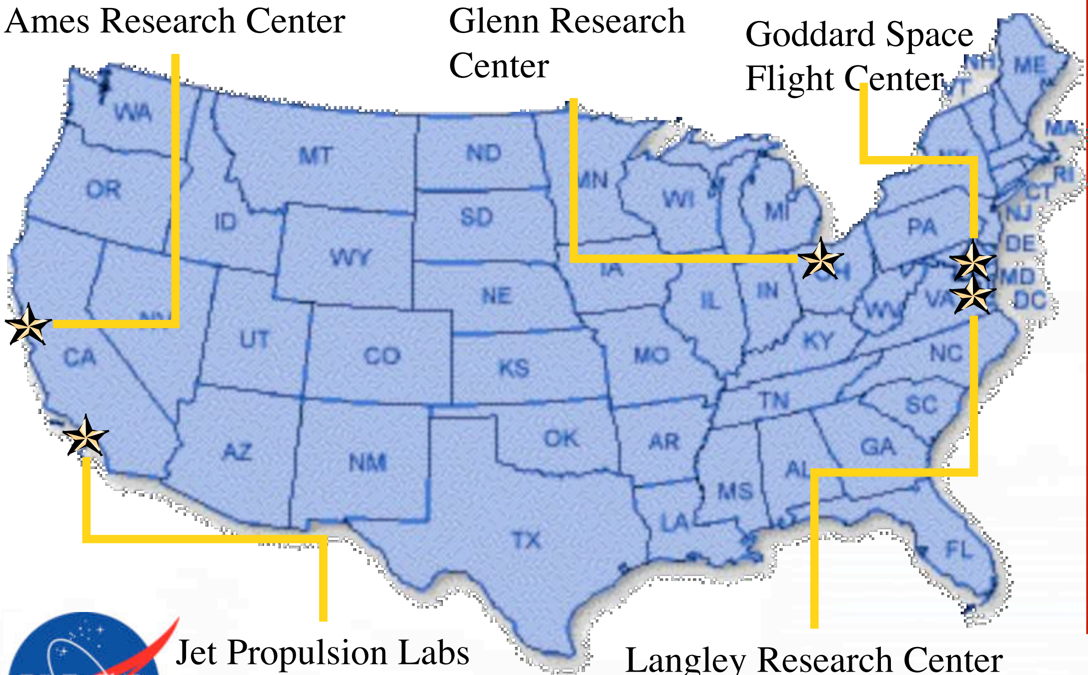
NAS
NASA ADVANCED SUPERCOMPUTING

Information Power Grid

Vision:

To make the practice of large-scale science and engineering, as well as other widely distributed, data intensive NASA activities, much more effective than it is today.

Grid technology is the foundation to making this vision a success



- Grid Nodes
  1. 1024 CPU SGI O3K IRIX ARC
  2. 512 CPU SGI O3K IRIX ARC
  3. 512 CPU SGI Altrix LINUX ARC (in progress)
  4. 128 CPU (node) LINUX Cluster GRC
  5. 128 CPU SGI O2K IRIX ARC
  6. 128 CPU SGI O2K IRIX ARC
  7. 64 CPU SGI O2K IRIX ARC
  8. 32 CPU SGI O2K IRIX ARC
  9. 24 CPU SGI O3K IRIX ARC
  10. 24 CPU SGI O2K IRIX GRC
  11. 16 CPU SGI O2K IRIX LaRC
  12. 16 CPU SGI O2K IRIX ARC
  13. 12 CPU (node) LINUX Cluster (in progress)
  14. 8 CPU SGI O3K IRIX LaRC
  15. 8 CPU SUN Ultra SPARC3 ARC - Storage
  16. 8 CPU SGI O2K IRIX ARC - Storage
  17. 8 CPU SGI O3K IRIX JPL
  18. 8 CPU SGI O3K IRIX GSFC (planned)
  19. 4 CPU LINUX ARC (planned)

NAS NASA ADVANCED SUPERCOMPUTING

Outline

- What are grid services and their importance to Earth Science
- How can grid services support Earth Science applications
- NASA's role in grids and grid services
- Roadmap to an Earth Science oriented service based cyberInfrastructure

Vision of Grid Use For Earth Sciences

- Goal: Implement an Earth Science Grid that integrates
  - Earth science data archives (EOSDIS/ESIP/REASoN)
  - Persistent Earth Science services that operate on Earth Science data
- Value:
  - Will provide an environment with re-usable, persistent services that are available to the whole Earth Science community
  - Will reduce the time needed to utilize Earth Science data
  - Could provide an outlet for making Earth Science research and development available to the whole Earth Science community as services.

Roadmap to an Earth Science Grid

- **Phase 1**: Provide grid access to the Earth Science archives (EOSDIS/ESIPS)
  - *Step 1.1*: provide grid access to disk resident data (data pools) associated with each archive. [Demonstrated in numerous ARC/GRC/CEOS work]
  - *Step 1.2*: Provide grid access to data resident on mass storage systems. [Demonstrated in June 03 CNIS ARC/LaRC grid milestone]
    - This capability exists with current technologies (such as the grid-enabled Storage Resource Broker)
    - Develop approach for mediating access since the existing data path out of mass storage systems are narrow.
  - *Step 1.3*: Provide grid accessible metadata catalogs to permit grid applications to automatically locate desired data based on data criteria
    - Access the data via the grid
    - Produce results whose metadata description is stored back in the metadata catalog.

    [Various examples within the grid community, part of current ARC CNIS milestone]

Roadmap to an Earth Science Grid (contd.)

- **Phase 2**: Provide grid services that can perform core Earth science processing
  - *Step 2.1*: Provide persistent services to perform common Earth Science functions such as subsetting, reformatting, data mining [Services are the focus of much of the global grid community work]
  - *Step 2.2*: Integrate archives and services into a seamless infrastructure that permits scientists to develop workflows
    - To extract desired data from archive
    - To move data through needed processing services
    - To send results to user or store results on archive as new data products.

Roadmap to an Earth Science Grid

- **Phase 3**: Extend the Earth Science cyberinfrastructure into space, making the various satellite-based sensors part of the cyberinfrastructure.
  - Step 3.1: Extend the cyberinfrastructure fabric into space, which under the current state-of-the art for grids will require that the Internet Protocols (e.g., IP) be extended into space and onto satellites and issues of reduced data rates, periodic disconnects and high latency be addressed.
  - Step 3.2: Implement sensors as grid services

Conclusions

- Grid technologies available currently can provide the basis for implementing an Earth Science Grid for early adopters.
- ARC stands ready to work with the Earth Science community to make this a reality.

Questions?

Backup Slides

Summary of Selected CEOS Projects

- USGS (US Geological Survey):
  - Explore use of GRID technologies for the delivery/reception of earth science data
- NOAA (National Oceanic and Atmospheric Administration):
  - Provide access to climate and numerical weather prediction models for analysis and intercomparison
  - Foster research to study complex earth systems using collections of distributed data
- ESA (European Space Agency):
  - Generic infrastructure to allow seamless plug-in of specific data handling & application services
  - Support on-demand user-driven data integration
- NASA GSFC (Goddard Space Flight Center) and GMU (George Mason University):
  - Integration of Grid and Open GIS Consortium (OGC) Web Services
- GMU (George Mason University):
  - Provide the ability to advertise and deliver virtual datasets
- UAH (U. of Alabama in Huntsville):
  - Compute-intensive data mining and machine learning applications in the Earth sciences
- DutchSpace and ESA/ESRIN (European Space Research Institute):
  - Simulators of EO instruments and data processing software working together using Computational Grid technology

NASA ADVANCED SUPERCOMPUTING